



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Evaluation of Host/Guest Binding Thermodynamics of Model Cavities with Grid Cell Theory

Citation for published version:

Michel, J, Henchman, RH, Gerogiokas, G, Southey, MWY, Mazanetz, MP & Law, R 2014, 'Evaluation of Host/Guest Binding Thermodynamics of Model Cavities with Grid Cell Theory', *Journal of Chemical Theory and Computation*, vol. 10, no. 9, pp. 4055–4068. <https://doi.org/10.1021/ct500368p>

Digital Object Identifier (DOI):

[10.1021/ct500368p](https://doi.org/10.1021/ct500368p)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Early version, also known as pre-print

Published In:

Journal of Chemical Theory and Computation

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Evaluation of Host/Guest Binding Thermodynamics of Model Cavities with Grid Cell Theory

Julien Michel,^{†} Richard H. Henchman,[‡] Georgios Gerogiokas,[†] Michelle W. Y. Southey,[§]
Michael P. Mazanetz[§] and Richard J. Law[§]*

[†] EastCHEM School of Chemistry, Joseph Black Building, The King's Buildings, Edinburgh,
EH9 3JJ, United Kingdom [‡] Manchester Institute of Biotechnology, The University of
Manchester, 131 Princess Street, Manchester M1 7DN, United Kingdom and School of
Chemistry, The University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom
[§] Evotec (UK) limited, Innovation Drive 114 Milton Park, Abingdon, Oxfordshire, OX14 4RZ,
United Kingdom

Keywords: hydration, thermodynamics, molecular recognition, cell theory

Abstract: A previously developed Cell Theory (CT) model of liquid water has been used to evaluate the excess thermodynamic properties of confined clusters of water molecules. The results are in good agreement with reference thermodynamic integration (TI) calculations, suggesting that the model is adequate to probe the thermodynamic properties of water at interfaces or in cavities. Next, the Grid Cell Theory (GCT) method has been applied to elucidate the thermodynamic signature of non-polar association for a range of idealized host/guest systems. Polarity and geometry of the host cavities were systematically varied and enthalpic and

entropic solvent components spatially resolved for detailed graphical analyses. Perturbations in the thermodynamic properties of water molecules upon guest binding are restricted to the immediate vicinity of the guest in solvent exposed cavities, whereas longer-ranged perturbations are observed in buried cavities. Depending on the polarity and geometry of the host, water displacement by a non-polar guest makes a small or large enthalpic or entropic contribution to the free energy of binding. Thus no assumptions about the thermodynamic signature of the hydrophobic effect can be made in general. Overall the results warrant further applications of GCT to more complex systems such as protein-ligand complexes.

1. Introduction

Many crucial molecular recognition events between organic or biomolecules occur in aqueous conditions and much effort has been directed at the development of molecular modelling methods to quantify the influence of water on the energetics of fundamental processes such as protein-ligand binding, solute solubility or membrane permeability.¹ This report focuses on the important contribution of water to the energetics of host/guest association. It is well appreciated that desolvation of hosts and guests in organic and biomolecular systems often yields a large contribution to the binding free energy,² or kinetics.³ In the area of protein-ligand association, particular attention has focused in recent years on the influence of binding site water molecules. The traditional view is that displacement of ordered binding site water molecules by adequate modifications of a ligand leads to gains in affinity by increasing solvent entropy. However, successful application of this principle is not guaranteed as it requires the design of an analog endowed with a suitable water-displacing moiety precisely positioned so as to offset the energetic penalty incurred by the loss of hydrogen-bonding interactions between the receptor and the displaced water molecule.⁴ Direct experimental evaluation of the free energy, enthalpy and

entropy of an ordered water molecule remains elusive and computational approaches present an attractive alternative. To this end different methodologies have been proposed, based for instance on inhomogeneous fluid solvation theory (IFST) (*e.g.* Watermap, STOW, GIST),⁵⁻⁹ semi-continuum electrostatics (SZMAP),¹⁰ 3D-RISM,¹¹⁻¹³ VISM,¹⁴ various free energy calculation protocols,¹⁵⁻¹⁹ and more rapid empirical scoring functions.²⁰ Recent findings from biophysical studies pursued with the help of such computational methods have led some to question the transferability of the hydrophobic effect for non-polar solvation to host/guest association because large enthalpic signatures have been observed upon ligand binding instead of the expected entropic contribution.²¹⁻²² There is controversy about the meaning and interpretation of these results because entropy-enthalpy compensating effects due to solvent reorganization or host/guest conformational changes may cloud the molecular driving forces of ligand binding.²³ Even for arguably simpler processes such as the hydration of small hydrophobic solutes, there is still controversy about the interpretation of the observed entropy loss.²⁴ These difficulties arise because although methods to estimate free energy are well established, it remains challenging to compute and interpret solvent entropies.²⁵

New methodologies are warranted to explore these issues and address current controversies. Recently our groups have proposed the Grid Cell Theory (GCT) methodology to spatially resolve the free energy, enthalpy and entropy of water in the vicinity of small organic molecules.²⁶ In this approach the cell theory method (CT) is used to compute solvent entropies by post-processing of a molecular dynamics simulation.²⁷ Hydration free energies, enthalpies and entropies for small organic molecules predicted with this approach were found to be in good agreement with experimental data and reference thermodynamic integration calculations. However, before the approach is routinely applied to more complex solute surfaces, such as that

of a cavity in a protein, it is important to assess whether the CT method reproduces the entropy of water in environments that deviate from bulk-like conditions. The present report addresses this issue with the use of idealized cavities and host/guest systems that enable well converged predictions of thermodynamic quantities and unambiguous comparison to benchmarks. First, the excess free energy, enthalpy and entropy of clusters of water molecules are computed under various degrees of confinement and compared with reference thermodynamic integration calculations. Second, the thermodynamic properties of interfacial water molecules are computed and spatially resolved for a range of cavities varying in geometry and polarity. The results extend findings from recent simulation studies of hydrophobic interactions with model enclosures and plates,²⁸⁻³¹ and illustrate a remarkable dependence of the thermodynamic signature of guest binding on host cavity water properties which has strong implications for rational drug design scenarios.

2. Theory and Methods

2.1 Excess thermodynamic properties of confined clusters of water molecules

The principle objective is to derive a protocol to rigorously compare excess thermodynamic properties computed with cell theory (CT) or thermodynamic integration (TI) methodologies. CT directly yields a partition function whereas TI estimates ratios of partition functions. Thus comparison requires the definition of a reference state accessible by the two techniques.

Excess properties of a confined cluster of N non-interacting water molecules

The expression of the classical canonical partition function is given by eq 1.

$$(1) Q_N = \frac{1}{N!h^{3N}} \int \dots \int e^{-\beta H(\mathbf{p}, \mathbf{r})} d\mathbf{p} d\mathbf{r}$$

Under the rigid-rotor approximation, the Hamiltonian of a solution of N rigid non-interacting water molecules is given by eq 2.³²

$$(2) H(\mathbf{p}, \mathbf{r}) = \frac{1}{2m} \sum_{i=1}^N (p_{x_i}^2 + p_{y_i}^2 + p_{z_i}^2) + \sum_{i=1}^N \left\{ \frac{\sin^2 \gamma_i}{2I_A} \left[p_{\beta_i} - \frac{\cos \gamma_i}{\sin \beta_i \sin \gamma_i} (p_{\alpha_i} - \cos \beta_i p_{\gamma_i}) \right]^2 + \right. \\ \left. \frac{\cos^2 \gamma_i}{2I_B} \left[p_{\beta_i} + \frac{\sin \gamma_i}{\cos \beta_i \cos \gamma_i} (p_{\alpha_i} - \cos \beta_i p_{\gamma_i}) \right]^2 + \frac{1}{2I_C} p_{\gamma_i}^2 \right\} + \\ U(x_1, y_1, z_1, \alpha_1, \beta_1, \gamma_1, \dots, x_N, y_N, z_N, \alpha_N, \beta_N, \gamma_N)$$

where the $3N$ atomic coordinates have been defined using rigid body representations. Specifically, x_i, y_i, z_i is the position of the oxygen atom of water molecule i , and $\alpha_i, \beta_i, \gamma_i$ are the Euler angles that define the orientation of water molecule i in its principal frame of reference. I_A, I_B, I_C are the principal moment of inertia and m the mass of a water molecule. $p_{x_i}, p_{y_i}, p_{z_i}, p_{\alpha_i}, p_{\beta_i}, p_{\gamma_i}$ are the linear and angular momenta of water molecule i . U is the potential energy function. As usual, integration over the momenta is possible, giving eq 3.

$$(3) Q_N = \frac{1}{N!} \left(\frac{2\pi m}{\beta h^2} \right)^{\frac{3N}{2}} \frac{8\pi^2}{\sigma} \left(\frac{2\pi}{\beta h^2} \right)^{\frac{3N}{2}} (I_A I_B I_C)^{\frac{N}{2}} Z_N$$

where σ is the symmetry number of a water molecule, $\beta = (1/k_B T)$, k_B is the Boltzmann constant, T is the temperature, h is the Planck constant and Z_N is the configuration integral. Each water molecule is restrained by a flat-bottom harmonic potential, and since there is no coupling between particles, the potential energy is given by eq 4.

$$(4) U(x_1, y_1, z_1, \dots, \alpha_N, \beta_N, \gamma_N) = \sum_{i=1}^N U(x_i, y_i, z_i, \alpha_i, \beta_i, \gamma_i)$$

where the potential energy of a single water molecule is given by eq 5.

$$(5) U(x_i, y_i, z_i, \alpha_i, \beta_i, \gamma_i) = 0 \text{ if } \|\mathbf{r}_i - \mathbf{r}_r\| \leq D \\ = K(\|\mathbf{r}_i - \mathbf{r}_r\| - D)^2 \text{ if } \|\mathbf{r}_i - \mathbf{r}_r\| > D$$

where K is the force constant of the restraining potential, \mathbf{r}_r the equilibrium position vector of the restraining potential, and D the flat-bottom radius. The configuration integral for one confined non-interacting water molecule is given by eq 6.

$$(6) Z_{FB} = \iiint_{-\infty}^{+\infty} \int_0^{2\pi} \int_0^\pi \int_0^{2\pi} e^{-\beta U(x_i, y_i, z_i, \alpha_i, \beta_i, \gamma_i)} dx_i dy_i dz_i d\alpha_i d\beta_i d\gamma_i$$

The potential energy defined by eqs 4,5 has no dependence on the orientation of the water molecule, thus it is possible to separate translational and rotational components, $Z_{FB} = Z_{tr,FB} Z_{rot,FB}$, where the components are given by eq 7 and 8.

$$(7) Z_{tr,FB} = \iiint_{-\infty}^{+\infty} e^{-\beta U(x_i, y_i, z_i)} dx_i dy_i dz_i$$

$$(8) Z_{rot,FB} = \int_0^{2\pi} d\alpha_i \int_0^\pi \sin \beta_i d\beta_i \int_0^{2\pi} d\gamma_i = 8\pi^2 \text{rad}^2$$

The excess free energy ΔF_{FB} , internal energy ΔU_{FB} and temperature-entropy product $-T\Delta S_{FB}$ follow from the standard expressions.³²

$$(9) \Delta F_{FB} = -\frac{1}{\beta} \ln \frac{Z_{tr,FB}}{V_{id}}$$

$$(10) \Delta U_{FB} = \iiint_{-\infty}^{+\infty} \frac{U(x_i, y_i, z_i, \alpha_i, \beta_i, \gamma_i) e^{-\beta U(x_i, y_i, z_i, \alpha_i, \beta_i, \gamma_i)}}{Z_{tr,FB}} dx_i dy_i dz_i$$

$$(11) -T\Delta S_{FB} = -\frac{1}{\beta} \ln \frac{Z_{tr,FB}}{V_{id}} + \Delta U_{FB}$$

where a reference ideal-gas volume V_{id} has been defined. For the process considered here the exact value is not important and we set $V_{id} = 1 \text{ \AA}^3$. Eq 7 and 10 can be conveniently evaluated accurately using numerical integration. For instance with $T = 298.0 \text{ K}$, $K = 10.0 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ and $D = 3.5 \text{ \AA}$, $Z_{tr,FB} = 215.476 \text{ \AA}^3$, $\Delta F_{FB} = -5.768 \text{ kcal}\cdot\text{mol}^{-1}$, $\Delta U_{FB} = 0.053 \text{ kcal}\cdot\text{mol}^{-1}$, $-T\Delta S_{FB} = -5.821 \text{ kcal}\cdot\text{mol}^{-1}$. Again, since the potential energy of each water molecule is independent from the position and orientations of other water molecules, $Z_N = (Z_{FB})^N$, and it follows that $\Delta F_{FB}^N = N\Delta F_{FB}$, $\Delta U_{FB}^N = N\Delta U_{FB}$, $-T\Delta S_{FB}^N = -NT\Delta S_{FB}$.

Excess properties of a mixture of N non-interacting and M interacting confined water molecules

Let the confined cluster of N non-interacting water molecules now contain M additional rigid water molecules that interact with each other through a standard pairwise additive classical potential energy function. The configuration integral is:

$$(12) Z_{N,M} = (Z_{FB})^N Z_M$$

where Z_M is a $6M$ dimensional integral that is not analytically soluble for $M > 1$ (for $M=1$, $Z_{N,1} = Z_{N+1,0}$). For $M > 1$ the non-interacting and interacting water molecules are also distinguishable and this would give rise to a $N!M!/(N+M)!$ term when subtracting the ideal partition function. These terms will not be considered further as they do not affect the comparison of excess quantities computed by cell theory or thermodynamic integration. The cell theory formalism of liquid water proposed by Henschman is used to approximate Z_M with eq 13.²⁷

$$(13) Z_M = (Z_{CT,M})^M$$

where the subscript CT,M indicates that the $6M$ configuration integral is estimated using a 6 dimensional effective configuration integral. To compute the effective configuration integral, each of the M water molecules is assumed to be described by an effective potential $\varphi(\mathbf{r}_i)$ that is independent of the positions and orientations of other water molecules and defined by eq 14.

$$(14) \varphi(\mathbf{r}_i) = \varphi(x, y, z, \theta_x, \theta_y, \theta_z) = \varphi_0 + \frac{1}{2}k_x x^2 + \frac{1}{2}k_y y^2 + \frac{1}{2}k_z z^2 + \frac{1}{2}k_{\theta_x} \theta_x^2 + \frac{1}{2}k_{\theta_y} \theta_y^2 + \frac{1}{2}k_{\theta_z} \theta_z^2$$

where x, y, z measure translations along the principal axes of water molecule i and $\theta_x, \theta_y, \theta_z$ measure rotations around the principal axes. $k_x, k_y, k_z, k_{\theta_x}, k_{\theta_y}, k_{\theta_z}$ are the force constants of the uncoupled one dimensional oscillator/pendulum describing translational/rotational motions along each dimension. φ_0 is the energy minimum. The configuration integral is given by eq 15.

(15)

 $Z_{CT,M}$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_0^{2\pi} \int_0^{2\pi} \int_0^{2\pi} \Omega_M e^{-\beta \left[\varphi_0 + \frac{1}{2} k_x x^2 + \frac{1}{2} k_y y^2 + \frac{1}{2} k_z z^2 + \frac{1}{2} k_{\theta_x} \theta_x^2 + \frac{1}{2} k_{\theta_y} \theta_y^2 + \frac{1}{2} k_{\theta_z} \theta_z^2 \right]} dx dy dz d\theta_x d\theta_y d\theta_z$$

where Ω_M is the orientational number of water (see below). The small-angle approximation is used to treat the rotations around the three axes as harmonic oscillations, which let eq 15 be rewritten as:

(16)

$$Z_{CT,M} = \Omega_M e^{-\beta \varphi_0} \prod_{i=x,y,z} \int_{-\infty}^{+\infty} e^{-\frac{\beta}{2} k_i i^2} di \int_{-\infty}^{+\infty} e^{-\frac{\beta}{2} k_{\theta_i} \theta_i^2} d\theta_i = \Omega_M e^{-\beta \varphi_0} \prod_{i=x,y,z} \sqrt{\frac{2\pi}{\beta k_i}} \sqrt{\frac{2\pi}{\beta k_{\theta_i}}}$$

and since for a one dimensional harmonic oscillator $F_i = \sqrt{\frac{2k_i}{\beta\pi}}$ and $\tau_i = \sqrt{\frac{2k_{\theta_i}}{\beta\pi}}$,²⁷ the expression simplifies to:

$$(17) Z_{CT,M} = e^{-\beta \varphi_0} \left(\frac{2}{\beta} \right)^6 \frac{\Omega_M}{\prod_{i=x,y,z} F_i \tau_i}$$

where F_i and τ_i denote force/torque constants. Subtracting the logarithm of the ideal partition function for one water molecule with volume V_{id} , from the logarithm of the cell theory effective partition function, multiplying by $-\beta^{-1}$, and using eq 17, leads to eq 18:

$$(18) -\frac{1}{\beta} [\ln Q_{CT,M} - \ln Q_{id}] = -\frac{1}{\beta} \ln \frac{Z_{CT,M}}{8\pi^2 V_{id}} = \varphi_0 - \frac{1}{\beta} \ln \left[\left(\frac{2}{\beta} \right)^6 \frac{\Omega_M}{8\pi^2 V_{id} \prod_{i=x,y,z} F_i \tau_i} \right]$$

and by identification eq 19:

$$(19) \Delta F_{CT,M} = \Delta U_{CT,M} - T \Delta S_{CT,M}$$

For some analyses it is convenient to break down the excess entropy in eq 19 in vibrational, librational and orientational components.

$$\begin{aligned}
(20) -T\Delta S_{CT,M} &= -\frac{1}{\beta} \ln \left[\left(\frac{2}{\beta} \right)^3 \frac{1}{V_{id} \prod_{i=x,y,z} F_i} \right] - \frac{1}{\beta} \ln \left[\left(\frac{2}{\beta} \right)^3 \frac{1}{8\pi^2 \prod_{i=x,y,z} \tau_i} \right] - \frac{1}{\beta} \ln[\Omega_M] \\
&= -T\Delta S_{CT,M}^{vib} - T\Delta S_{CT,M}^{lib} - T\Delta S_{CT,M}^{ori}
\end{aligned}$$

The vibrational and librational terms in eq 20 measure the excess entropy arising from translations and rotations of the water molecules. The orientational term in eq 20 measures the excess entropy arising from the number of equivalent minimas in solution. In eqs 17,18,20 F_x , F_y , F_z , τ_x , τ_y , τ_z are the average magnitude of the forces and torques constants measured in the principal frame of reference of each of the M water molecules and Ω_M is the average orientational number. φ_0 is taken as the average potential energy of an interacting water molecule (i.e. $\varphi_0 = \langle U_M \rangle / M$). Contributions from pairwise terms are halved to avoid double counting. Eq 17 is only valid in the classical limit and breaks down for low forces and torques. For the librational term, an upper value of $8\pi^2 \text{ rad}^2$ for $\left(\frac{2}{\beta} \right)^3 \frac{1}{\prod_{i=x,y,z} \tau_i}$ is enforced. The orientational number Ω_M essentially accounts for the number of ways a water molecules can form equivalent hydrogen-bonding networks.²⁴ It is based on a generalisation of the Pauling's ice entropy model.³³

$$(21) \Omega_M = \max\left(1.0, \frac{(N_w-1)(N_w-1)^2}{2N_w}\right)$$

where N_w is the average number of neighboring water molecules in the coordination sphere of each water molecule, a cutoff of 3.4 Å is used by default. More elaborate treatments have been proposed elsewhere,³⁴ but are not considered in this study. Because the effective partition function is additive, the excess thermodynamic properties of a system made of M interacting water molecules and N non-interacting water molecules is given by:

$$(22) \Delta F_{excess}^{N,M} = N\Delta F_{FB} + M\Delta F_{CT,M}$$

$$(23) \Delta U_{excess}^{N,M} = N\Delta U_{FB} + M\Delta U_{CT,M}$$

$$(24) -T\Delta S_{excess}^{N,M} = -NT\Delta S_{FB} - MT\Delta S_{CT,M}$$

Alternatively, expressing excess quantities as a sum of differences between successive M values one can write:

$$(25) \Delta F_{excess}^{N,M} = (N + M) \Delta F_{FB} + \sum_{k=0}^{M-1} \Delta F_{M=k \rightarrow k+1}$$

$$(26) \Delta U_{excess}^{N,M} = (N + M) \Delta U_{FB} + \sum_{k=0}^{M-1} \Delta U_{M=k \rightarrow k+1}$$

$$(27) -T\Delta S_{excess}^{N,M} = \Delta F_{excess}^{N,M} - \Delta U_{excess}^{N,M}$$

where:

$$(28) \Delta F_{M=k \rightarrow k+1} = -\frac{1}{\beta} \ln \frac{(Z_{FB})^{N+M-k-1} Z_{k+1}}{(Z_{FB})^{N+M-k} Z_k} = -\frac{1}{\beta} \ln \frac{Z_{k+1}}{Z_{FB} Z_k}$$

$$(29) \Delta U_{M=k \rightarrow k+1} = \langle U_{k+1} \rangle - \langle U_k \rangle$$

Equation 28 can be solved using for instance using TI as shown in Eq 30.

$$(30) \Delta F_{M=k \rightarrow k+1} = \int_{\lambda=0}^{\lambda=1} \frac{\partial F(x_1, y_1, z_1, \dots, \alpha_M, \beta_M, \gamma_M, \lambda)}{\partial \lambda} d\lambda = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial U(x_1, y_1, z_1, \dots, \alpha_M, \beta_M, \gamma_M, \lambda)}{\partial \lambda} \right\rangle d\lambda$$

where λ is a coupling parameter that smoothly turns on the intermolecular interactions of one of the M water molecules. Eqs 22-27 provide a rigorous way to compare excess properties computed by CT or TI. In practice, computer simulations are used to numerically evaluate F_x , F_y , F_z , τ_x , τ_y , τ_z , Ω_M , $\langle U_M \rangle$, $\left\langle \frac{\partial U(x_1, y_1, z_1, \dots, \alpha_M, \beta_M, \gamma_M, \lambda)}{\partial \lambda} \right\rangle$ and finite-sampling errors limit the range of values for the radius D of the restraining potential, the number of interacting water molecules M and the temperature T .

2.2 Grid Cell Theory analyses of host-guest binding thermodynamics.

The Grid Cell Theory is an alternative formulation of the Cell Theory method where a volume of space is discretized into N_s volume elements of volume $V(k)$ that each contain a variable $N_w(k)$

number of water molecules. Details of the methodology have been presented elsewhere.²⁶

Briefly, cell parameters $\varphi_0(k)$, $\Omega_M(k)$, $F_{i=x,y,z}(k)$, $\tau_{i=x,y,z}(k)$ are computed for each voxel k .

The cell parameters of an arbitrary region s are then computed by weighting the cell parameters of the voxels belonging to s , i.e.

$$(31) \varphi_0(s) = \frac{1}{N_w(s)} \sum_{k=1}^{N_s} I(s) N_w(k) \varphi_0(k)$$

$$(32) \Omega_M(s) = \max(1, \frac{1}{N_w(s)} \sum_{k=1}^{N_s} I(s) N_w(k) \Omega_M(k))$$

$$(33) F_i(s) = \frac{1}{N_w(s)} \sum_{k=1}^{N_s} I(s) N_w(k) F_i(k)$$

$$(34) \tau_i(s) = \frac{1}{N_w(s)} \sum_{k=1}^{N_s} I(s) N_w(k) \tau_i(k)$$

where $I(s)$ is an indicator function that is equal to 1 if voxel k belongs to s , 0 otherwise. $N_w(s)$ is the average number of water molecules within region s . Eqs. 31-34 enable the computation of the effective partition function and excess properties for s according to eqs 18-19. For host/guest analyses, it is convenient to express excess water properties of s relative to bulk conditions. Additionally, neglecting contributions from pressure-volume terms, we write:

$$(35) \Delta H_w(s) = N_w(s) \Delta H_w^n(s) = N_w(s) [\Delta U_{(s)}^{CT} - \Delta U_l^{CT}]$$

$$(36) -T\Delta S_w(s) = -N_w(s) T\Delta S_w^n(s) = N_w(s) [-T\Delta S_{(s)}^{CT} + T\Delta S_l^{CT}]$$

where the subscript l indicates quantities computed using reference bulk and ideal parameters. The superscript n indicates normalized (per water) quantities. The above two quantities can be summed to estimate a water free energy contribution $\Delta G_w(s)$. For some analyses, it is also convenient to break down $-T\Delta S_w(s)$ in vibrational, librational and orientational components, $-T\Delta S_w^{vib}(s)$, $-T\Delta S_w^{lib}(s)$, $-T\Delta S_w^{ori}(s)$ respectively. To establish enthalpy and entropy changes for the reversible association of a guest G with a host H one must analyse water properties over

three separate regions **s1**, **s2**, **s3** as depicted in Figure 1. Under the assumption that the host and guest are rigid, the enthalpies and entropies of binding are given by eqs 37-38:

$$(37) \Delta H_{HG(\varphi)}^{bind} = \Delta E_{\varphi} + \Delta H_{w,\varphi}(\mathbf{s1}) - \Delta H_w(\mathbf{s2}) - \Delta H_w(\mathbf{s3})$$

$$(38) -T\Delta S_{HG(\varphi)}^{bind,\circ} = -T\Delta S_{G,\varphi}^{\circ} - T\Delta S_{w,\varphi}(\mathbf{s1}) + T\Delta S_w(\mathbf{s2}) + T\Delta S_w(\mathbf{s3})$$

where the first term in Eq 37 is the interaction energy of *H* with *G* in a fixed relative orientation φ . The regions **s1**, **s2**, **s3** are assumed to be sufficiently large such that bulk like properties for water are recovered beyond the edges of each region, and therefore have to be adjusted for the particular host/guest systems under consideration. The host and guest are assumed to be rigid and there is therefore no entropic contribution from changes in internal degrees of freedoms. The entropy of binding also requires the evaluation of changes in translational and rotational entropy for *H* and *G*. The number of translational and rotational minima of the bound guest is assumed to be one. Additionally, the host is assumed to exhibit negligible changes in translational/rotational entropy upon guest binding. This leaves only the contribution of the change in translational/rotational entropy for the guest to the entropy of binding. A system-frame cell model is used to estimate this contribution with Eq 39.³⁵

$$(39) -T\Delta S_{G,\varphi}^{\circ} = -\frac{1}{\beta} \ln \left[\frac{\sigma_G V_w^2}{V^{\circ} 8\pi^2} \prod_{i=x,y,z} \frac{1}{r_G^i} \frac{F_i(G)\tau_i(G)}{F_i(G,\varphi)\tau_i(G,\varphi)} \right]$$

where F_i , τ_i indicates half magnitudes of forces and torques components along the principal axes of the guest, measured in solution or in complex with the host in relative orientation φ , r_G^i is the radius of the guest along principal axis *i* to the edge of its van der Waals surface, V_w is the volume available to a water molecule in solution, V° is the volume available to the guest at the 1M standard-state concentration and σ_G is the symmetry number of the guest. Finally, eqs 37 and 38 can be summed to obtain the free energy of binding $\Delta G_{HG(\varphi)}^{bind,\circ}$.

2.3 Preparation of molecular models.

The TIP4P-Ew water model was used throughout.³⁶ A cluster of 12 water molecules was prepared using the program leap from the AMBER 11 software suite.³⁷ For the host/guest simulations, hosts were constructed by layering atomic sites with a spacing between adjacent sites of 4.19 Å in a rectangular region of minimum coordinates $x = 0$ Å, $y = 0$ Å, $z = 0$ Å and maximum coordinates $x = 20.93$ Å, $y = 25.12$ Å, $z = 20.93$ Å. The initial spacing was chosen such that the Lennard-Jones interaction energy between neighboring host sites is close to a minimum with the parameters used (see below). A solvent accessible cavity was then carved out by removing a variable number of sites and in the case of the complexes, inserting guest sites at the bottom of the host cavity. The host was aligned such that sites at the bottom of the cavity have a lower x coordinate than at the top of the cavity. The non-bonded Lennard-Jones parameters of the host sites were adapted from OPLS-UA parameters for methane ($r^* = 2.093$ Å and $\varepsilon = 0.294$ kcal•mol⁻¹).³⁸ The guest site Lennard-Jones parameters were taken as the host site parameters scaled down by 10% ($r^* = 1.884$ Å and $\varepsilon = 0.264$ kcal•mol⁻¹) so that the guest wouldn't fill completely the cavity. Adjacent guest sites were bonded with a harmonic potential of parameters $r_{eq} = 3.768$ Å and $K_{eq} = 500$ kcal•mol⁻¹•Å⁻². Host sites at the edges of the cavity were next converted into polar sites bearing a positive or negative partial charge q . Several different models of hosts and host/guest complexes were prepared, varying the polarity (by changing the magnitude of the charges $|q|$) and the geometry (by removing additional host sites along the x , y or z axes), see Figure 2 for an illustrative example. Each model host, guest or complex was solvated in a box of TIP4P-Ew water molecules extending 12 Å from the edge of the solute(s) using the program leap from the AMBER 11 software suite. The resulting models

were energy minimized and equilibrated under NPT conditions at 1 atm and 298 K (host/guest) or in non-periodic conditions (water clusters) using the software sander. A velocity-Verlet integrator and a time step of 2 fs was used. Temperature control was achieved with a Langevin thermostat and coupling constant of 5 ps^{-1} ,³⁹ whereas pressure control was achieved with a Berendsen barostat and coupling constant of 2 ps^{-1} .⁴⁰ The SHAKE method with a tolerance of 0.00001 \AA was applied to constrain intramolecular degrees of freedoms in water molecules and bonds involving hydrogen atoms in solutes.⁴¹⁻⁴² Host/guest sites were restrained to their input positions during the equilibration using harmonic positional restraints of $10000 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ and constraints applied to keep water molecules rigid. Electrostatic interactions were handled with the particle mesh Ewald method,⁴³ and a cutoff of 10 \AA . Lennard-Jones interactions were truncated after 10 \AA .

2.4 Molecular Dynamics simulations.

Production molecular simulations were performed with the software Sire/OpenMM. In the present study, this program results from the runtime linking of the general purpose molecular simulation package Sire revision 2014 (host/guest simulations) or revision 2373 (water cluster simulations),⁴⁴ with the GPU Molecular dynamics library OpenMM 5.1.⁴⁵ The water cluster simulations were performed at 298 K with a cutoff of 99 \AA and an atom-based Barker Watts reaction field with dielectric constant set to one.⁴⁶ The host/guest simulations were performed at 1 atm and 298 K using a reaction field non-bonded cutoff of 10 \AA and dielectric constant of 78.3 for the electrostatic interactions, and an atom based non-bonded cutoff of 10 \AA for the Lennard-Jones interactions. Host/guest sites were kept frozen during the dynamics. A velocity-Verlet integrator was used with a timestep of 2 fs. Temperature control was achieved with an Andersen

thermostat and a coupling constant of 10 ps^{-1} .⁴⁷ Pressure control was achieved by attempting isotropic box edge scaling Monte Carlo moves every 25 time-steps. The intramolecular degrees of freedom of water molecules were constrained using the OpenMM default tolerance settings.

Each water cluster was simulated for 22 ns, and the first 2 ns discarded to allow for equilibration. Unless otherwise stated, each host/guest system was simulated for 10 ns and the first 1 ns was discarded to allow for equilibration. Snapshots were stored every 1 ps in a DCD file format for subsequent analyses, thus a total of 9,000 snapshots were post-processed for a typical host/guest simulation. All molecular dynamics simulations were executed using the OpenCL platform of OpenMM on a cluster of Tesla-M2090 nodes.

2.5 Nautilus analyses.

Cell theory parameters for the water clusters were computed using the trajectory analysis software Nautilus.²⁶ For the host/guest complexes, grid cell theory analyses were performed, using cell theory bulk parameters for TIP4P-Ew water taken from a previous GCT study (Table 1 in ref²⁶). Additionally, the Nautilus implementation was extended to enable computation of enthalpies and entropies of binding as given by eqs 37-39. The symmetry number for the guest was taken as $\sigma_G = 4$. Eq 39 does not account for the effect of the positional constraints that were applied to host and guest sites during the simulations performed in this study. The change in guest translational/rotational entropy was therefore calculated with the additional approximation that the ratio of the products of forces and torques is equal to one. A previous cell-theory study of a set of protein-ligand complexes has shown that these terms contribute little to the guest entropy loss.³⁵

Efforts are underway to release Nautilus as standalone software. Briefly, Nautilus analyses entail three major steps: 1) Conversion of a molecular dynamics trajectory into a collection of solvent and guest cell files. 2) Conversion of solvent cell files into grid files. 3) Evaluation of the thermodynamic properties of water and the guest over a user defined region. For the graphical analyses of the host/guest simulations (region *s1* in Figure 1) the grid center was positioned on the average coordinate of the host sites (ca. $x=23.5 \text{ \AA}$, $y=25 \text{ \AA}$, $z=23.8 \text{ \AA}$) and the grid covered a rectangular region extending $\pm 10 \text{ \AA}$ in the y and z coordinates and $+25 \text{ \AA}$ and -10 \AA in the x coordinate. The same region was used for the host only simulations (region *s2*). For the guest simulations (region *s3*) a cubic region was centered on the average coordinates of the center of geometry of the guest ($x=17.8 \text{ \AA}$, $y=18.2 \text{ \AA}$, $z=18.3 \text{ \AA}$) and extended by $\pm 12 \text{ \AA}$ in the x , y and z coordinates. A grid spacing of 0.5 \AA was used throughout. For quantitative analyses and evaluation of thermodynamic signatures a sub-region of *s1* and *s2* was used by only considering grid points that belong to the interval $x \in [5, 55] \text{ \AA}$, $y \in [5, 35] \text{ \AA}$, $z \in [10, 35] \text{ \AA}$.

2.6 Thermodynamic integration calculations

Thermodynamic integration as implemented in Sire/OpenMM was used to estimate eq 25. Details of the potential energy function were exactly as used for the CT calculations. Stepwise decoupling of one water molecule from the cluster was performed in two stages, electrostatic interactions were removed first, followed by Lennard-Jones interactions. For each stage 24 λ values spanning the interval $[0.0, \dots, 1.0]$ were used initially. Free energy gradients at each λ value were evaluated using a finite-difference thermodynamic integration (FDTI) approach with $\Delta\lambda$ set to 0.001.⁴⁸ In order to avoid numerical instabilities, soft-core potential energy functions were used for transformations of the Lennard-Jones parameters.⁴⁹ An implementation identical to

Michel et al. was used and the softening parameters were set to $n=0$ and $\delta=3.0$ throughout.⁵⁰ Free energy changes were then obtained by numerical integration of the free energy gradients using a polynomial regression scheme.⁵¹ In some instances ($M=12$, $D=3.0$ Å and $M=12$, $D=2.75$ Å) sharp variations in free energy gradients were observed in the Lennard-Jones decoupling step and calculations at up to four additional λ values were performed to reduce numerical integration errors. Each λ value was simulated for 1.1 ns, with the first 100 ps discarded to enable equilibration. Changes in enthalpy were estimated from eq 29 by extending end-states simulations (i.e. $\lambda=0.0$, electrostatic decoupling step, and $\lambda=1.0$, Lennard-jones decoupling step) to 20 ns to obtain well converged total average potential energies.

3. Results and Discussion

3.1 Confined water clusters

Figure 3 compares the excess free energies, internal energies and entropies computed for several water clusters under a range of different confinements with the CT and TI methods. Because both methods use the same approach to compute the internal energy, they should produce identical results for this quantity. In practice differences may be expected because the energies are averaged over different simulation trajectories of finite length. The relatively good agreement in the computed internal energies observed for a broad range of M values suggests that the results are well converged and that the computed entropies can be reliably compared. The only noticeable discrepancy occurs at low D and high M values (Figure 3D). This corresponds to situations where the density of interacting water molecules is high and water mobility decreases dramatically, causing slow convergence of the average potential energy or free energy gradients. Turning to the entropy, it is apparent that both theories predict a

significant loss of entropy upon turning on intermolecular interactions between water molecules; depending on the confining potential, the potential energy decreases by ca. 80-90 kcal•mol⁻¹ between $M=0$ to $M=12$, but this is largely offset by an entropy loss of ca. 60 kcal•mol⁻¹. There is an obvious discrepancy between CT and TI for $M=1$ no matter the confinement. This corresponds to a situation where a single interacting water molecule is only weakly constrained by the confining potential. As a result, very low forces and torques are measured during the simulation and the harmonic approximation of CT breaks down, causing an overestimation of the entropy of the water molecule. However, for $M=2$ and beyond the results are much closer to the TI data, suggesting that the harmonic approximation is valid for even weakly interacting clusters. Figure 3 also shows that the greater the confinement, the smaller the discrepancy in the entropies computed with the two methods. At low confinement (Figure 3 panels A & B) the discrepancy increases with the number of M interacting water molecules. However, at higher confinement (Figure 3 panels C & D) better agreement between CT and TI is observed at low and high M values. Over the interval $M=2$ -12 the average discrepancy in entropy per water molecule is 0.5 kcal•mol⁻¹ at low confinement ($D=5$ Å) and only 0.2 kcal•mol⁻¹ at low confinement ($D=2.75$ Å). Overall the results suggest that CT is a valid alternative to TI to predict the entropy of confined water molecules and the methodology should be useful for studies of the energetics of water molecules at complex interfaces such as host/guest cavities.

3.2 Model host-guest systems

Unlike protein-ligand complexes for which it is challenging to obtain converged estimates of solvent enthalpies and entropies, the present idealized host-guest systems offer the opportunity to systematically vary simulation parameters at a reasonable computing cost. Owing

to the axis of symmetry in the xy plane, convergence of the simulations can be visually assessed by inspection of the isocontours of the computed thermodynamic properties. A typical setup for a polar cavity is shown in Figure 2 which depicts isocontours of the free energy of water $\Delta G_w^n(\mathbf{s})$. The contours show that first-shell water molecules outside of the cavity are moderately destabilized with respect to bulk, whereas stabilizing regions are observed for water molecules in contact with polar cavity sites. Visualization of the computed trajectories shows that the most stable regions are observed between two negatively charged cavity sites at the bottom of the cavity where a water molecule can donate two hydrogen bonds and accept one from the host. The reverse motif where water accepts two hydrogen-bonds and donates one to the host is also observed in the simulations but is less stable (data not shown). These observations are in agreement with the asymmetric hydration of water by hydrogen-bond donors or acceptors.⁵²⁻⁵⁸ Water in the bottom central region of the cavity is less ordered but also interacts more weakly with the host and consequently hydration of this region is unfavorable (Figs 2A and 2B).

Figure 4 depicts the cavity water density relative to bulk as a function of the magnitude of the partial charges on the cavity polar sites. Below $|q| = 0.35$ e the cavity is largely dry, with only the upper part transiently hydrated. At $|q| = 0.35$ e the cavity undergoes multiple dewetting-wetting transitions on a ns time scale, as evidenced by the large standard deviation of the mean density. Bulk density is recovered for $|q| = 0.6$ e. Beyond this, visualization of the computed trajectories suggests that further increases in water density are observed at the expense of a slowing down of cavity water dynamics (data not shown). The behavior of water in these different cavities is reminiscent of the hydration of protein binding sites. The hydrophobic binding site of bovine β -lactoglobulin has been shown to be dry in *in vitro* conditions by NMR magnetic relaxation dispersion experiments and a range of computational methods.⁵⁹ Young et

al. have computationally identified several dry protein binding sites,⁶⁰ whereas Matthews and Liu have reviewed the experimental evidence for dry protein cavities and concluded in favor of their occurrence for at least small non-polar cavities that cannot accommodate more than one water molecule.⁶¹ The binding site of the protein MUP-1 and COX-2, both fairly hydrophobic, have been shown to be sub-hydrated or even dry in *in vitro* conditions.^{8, 17, 62}

3.3 Influence of cavity polarity on water properties

The stability of interfacial water molecules was further assessed by performing Nautilus analyses of host cavity hydration in the dry-regime ($|q| = 0.25$ e), dry-wet regime ($|q| = 0.35$ e), and wet regime ($|q| = 0.45$ e). Figure 5 shows a 2D projection of the normalized excess thermodynamic properties of water and relative water density ρ_s ($\rho_s = N_w(s)/N_l(s)$ where the denominator is the number of bulk water molecules expected in a volume of space $V(s)$). At low polarity (Figure 5, top row) the cavity remains dry throughout the simulation. Depletion in water density immediately above the cavity is apparent and so are a structured first and second hydration shells. The water molecules in the first hydration shell of the host have a slightly less favorable enthalpy. Perturbations in water entropy are weaker in magnitude but extend further, with weak discernible patterns in the second hydration shell. Bulk-like behavior appears to be recovered at distances greater than ca. 4 Å from the host surface. At intermediate polarity (Figure 5, second row) the cavity undergoes frequent dewetting transitions on a nanosecond timescale. Two types of water behavior can be discerned. When hydrated, the top of the cavity (x ca. 22 Å) is occupied by weakly enthalpically and entropically destabilized water molecules. On the other hand the middle of the cavity (x ca. 18 Å) is occupied by water molecules that are enthalpically and entropically stabilized with respect to bulk. At higher polarity (Figure 5, third row) the

cavity is filled with a network of structured water molecules that are enthalpically stabilized and entropically destabilized with respect to bulk conditions. A significant contribution from the first hydration shell is also apparent. Upon guest binding all water molecules are displaced from the cavity (Figure 5, fourth row). Similar results are observed when the guest is bound for other $|q|$ values (data not shown). Water molecules in the first hydration shell of the bound guest are weakly stabilized at higher cavity polarities due to electrostatic interactions but this effect is not strong and the energetics of cavity dehydration are largely driven by the displacement of the water molecules that would otherwise have steric clashes with the guest. Figure 6 shows the thermodynamic signature of host/guest binding as a function of the host cavity polarity. The uncertainty in the computed energetics was estimated by computing the standard deviation of the mean for 3 independent repeat simulations at $|q| = 0.45$ e. The results, $\sigma=0.5$ kcal•mol⁻¹ for the enthalpy of binding and $\sigma=0.2$ kcal•mol⁻¹ for the entropy of binding indicate that the changes in thermodynamic signature with $|q|$ seen in Figure 6 are significant. For $|q|$ values lower than 0.30 e there is little variation in the thermodynamic signature since the host cavity is largely dry. Association is favored by ca. -18 kcal•mol⁻¹ because of a favorable enthalpic contribution from guest desolvation (ca. -4 kcal•mol⁻¹), host desolvation (ca. -3 kcal•mol⁻¹), and the favorable host-guest interaction energies (-12.5 kcal•mol⁻¹). The binding entropy includes a favorable contribution for guest desolvation (ca. -3 kcal•mol⁻¹) that is compensated by the guest loss of translational and rotational entropy (+4.6 kcal•mol⁻¹). Between $|q| = 0.35$ to 0.45 e the number of cavity water molecules increases rapidly and the increased entropy gain due to cavity water displacement causes a favorable change in binding entropy of ca. -3.8 kcal•mol⁻¹. This, however, is offset by an enthalpic change of ca. +25 kcal•mol⁻¹ owing to loss of strong hydrogen-bonding interactions between host polar sites and cavity water molecules. For $|q|$ values above 0.45 e

there is a small additional entropic gain from water displacement but stronger enthalpic loss, thus causing host/guest association to become unfavorable. This underscores that strongly enthalpically stabilized cavity water molecules cannot be productively displaced by a guest unless new hydrogen-bonding interactions of similar strength are formed between the host and guest.

For $|q|$ values between 0.35 to 0.55 e the number of water molecules displaced from the monitored region upon guest binding ranges from ca. 8 to ca. 14. The associated favorable entropic contribution of cavity water displacement to the free energy of binding increases from about $0.0 \text{ kcal}\cdot\text{mol}^{-1}$ to $0.5 \text{ kcal}\cdot\text{mol}^{-1}$ per displaced water molecule. These estimates can be compared to literature values. Dunitz estimated on the basis of the entropy difference between liquid water and ice that the entropy gain for water displacement in proteins should be no more than $2 \text{ kcal}\cdot\text{mol}^{-1}$, and likely much less as protein binding site water molecules retain some mobility.⁶³ Verdonk et al. estimated a contribution of $0.47 \text{ kcal}\cdot\text{mol}^{-1}$ per displaced water molecule from fitting to experimental data.⁶⁴ Figure 7 depicts a component analysis of the changes in solvent entropy upon guest binding. In all cases there is an unfavorable contribution from changes in water vibrational entropy which indicates that cavity water molecules have greater translational motions than in bulk. Changes in water librational entropy also oppose guest binding, but this term approaches zero as the cavity polarity is increased. This indicates that water molecules in the more polar cavities are as restricted in their rotational motions as in bulk. Nevertheless a favorable entropic contribution to guest binding is observed overall because of a large favorable contribution from the changes in water orientational entropy. Therefore, in the present simulations, the favorable entropic gain due to release of cavity water molecules in bulk is due to an increase in the number of hydrogen-bonding partners that more than offsets a

decrease in translational/rotational motions. The results obtained here can be compared to those reported by Irudayam and Henchman in an earlier cell theory study.³⁵ The authors estimated that displacing a single water molecule from a hydrophobic cavity in the protein barnase would yield an essentially zero change in water entropy due to compensation between vibrational/librational and orientational entropy components. On the other hand displacement of a single water molecule from an hydrophilic cavity in the protein BPTI was predicted to yield a favorable entropic contribution of $0.9 \text{ kcal}\cdot\text{mol}^{-1}$ due to a large favorable orientational entropy component.³⁵ These values are somewhat lower than the favorable entropic contribution that has been predicted with IFST methods for displacing a water molecule from the binding sites of HIV-1 protease ($2.9 \text{ kcal}\cdot\text{mol}^{-1}$),⁶⁵ concavilin A ($2.0 \text{ kcal}\cdot\text{mol}^{-1}$),⁶⁶ or cyclophilin A (ca. $1.8 \text{ kcal}\cdot\text{mol}^{-1}$).⁶⁷ It has been suggested that neglect of correlation terms in the IFST implementation of Li and Lazaridis may explain the higher values obtained.³⁵ Additional detailed comparative studies of CT and IFST methods applied to identical systems are desirable to establish the range of water entropies at protein interfaces.

3.4 Influence of cavity width on water properties

The low water molecules occupancy of cavities under low polarity conditions ($|q| < 0.35 \text{ e}$) warrants further consideration. To determine whether the phenomenon was a consequence of the curvature of the cavity, simulations were repeated with enlarged cavities along the y and or z planes and at $|q|$ values near the dry-wet transition seen in Figure 4 ($|q| = 0.25 \text{ e}, 0.30 \text{ e}, 0.35 \text{ e}$). Figure 8 summarizes the main results from this analysis. Although the solvent-accessible volume of the enlarged cavities is four times greater than cavities studied previously, no significant wetting is observed at $|q| = 0.30 \text{ e}$ (Figure 8, first row). In contrast with the narrower apolar

cavities (Figure 5, first row), water molecules in the first hydration shell immediately above the enlarged cavity are more destabilized and there is a clear depletion in water density in this region (Figure 8, first row, x ca. 25 Å and $y \in [5,15]$ Å). The effect is largely enthalpic in origin and can be traced to the decreased solute-solvent interactions experienced by water molecules in this region. Additional polar sites placed at the base of the cavity (Figure 8, second row), or increasing the polarity of the sites ($|q| = 0.35$ e, Figure 8 third row), are both insufficient to wet the cavity. Consequently the water properties are very similar in all cases. However, placing additional sites and increasing polarity is sufficient to cause wetting of the cavity (Figure 8, fourth row). Under these conditions, bulk-like density is achieved near polar sites at the base of the cavity. Intriguingly water molecules located at the base of the cavity and near the corners (Figure 8, fourth row, x ca. 16 Å, y ca. 4 Å and x ca. 16 Å, y ca. 16 Å) are both enthalpically and entropically stabilized with respect to bulk. Low density water regions are observed in the middle of the cavity, and water molecules in this region are moderately destabilized with respect to bulk. No strong fluctuations in cavity water density are observed during the simulations, and significant regions of the cavity remain dry.

Overall the picture that emerges is that cavity wetting is less sensitive to the width of the cavity than to the polarity of the cavity. Thermodynamic signatures for guest binding were not computed for these enlarged cavities owing to uncertainties in the optimum placement of the guest, and difficulties in sampling adequately with molecular dynamics the solvent occupancy of residual pockets of space left between the host and the guest. Addressing these technical difficulties will require future implementations of specialized water-placement algorithms,¹⁷ or Grand-Canonical Monte Carlo sampling protocols.¹⁹

3.5 Influence of cavity depth on water properties

As an additional test, the magnitude of the host partial charges was kept constant ($|q| = 0.45\text{ e}$) but the depth of the host cavity was increased. Figure 9 depicts the results for three different cavities. Results for the x coordinate cavity base at 15 Å have already been shown (Figure 5, third row). As the x coordinate of the cavity base decreases from $x = 15$ to $x = 12\text{ Å}$ (Figure 9, top row) there is a shift in the position of the stable cavity water molecules so as to maintain hydrogen-bonding interactions with host polar sites. However, as the cavity waters do still exchange with bulk, this causes the formation of an additional layer of weakly stabilized water molecules near the top of the cavity (x ca. $22\text{-}25\text{ Å}$). Decreasing further the cavity base to $x = 9\text{ Å}$ (Figure 9, second row) also causes a shift in the position of the stabilized water molecules, but the water molecules in the intermediate hydrophobic region (x ca. $17\text{-}25\text{ Å}$) are now unstable with respect to bulk. Finally, decreasing further the cavity base to $x = 6\text{ Å}$ (Figure 9, third row) dramatically disrupts hydration, with only a few water molecules retained at the bottom of the cavity and almost complete drying of the rest of the cavity. Visualization of the computed trajectories shows that water diffusion in the cavity dramatically slows down and longer simulations (ca. 100 ns) are needed to observe a statistically significant number of water exchanges between the bottom of the cavity and bulk (data not shown).

Upon guest binding to the bottom of the deepest cavity complete drying is observed (Figure 9, fourth row). The same result is observed for all other cavities (data not shown). Thus adequate assessment of the changes in water energetics requires here consideration of perturbations in water network structure that extend up to about 10 Å away from the surface of the guest. Figure 10 depicts the thermodynamic signature of host/guest binding for different sized cavities. Increased burial of the guest in the host favors slightly binding to the deeper

cavities due to stronger host/guest interaction energy (from $-12.5 \text{ kcal}\cdot\text{mol}^{-1}$ at $x = 15 \text{ \AA}$ to $-14.0 \text{ kcal}\cdot\text{mol}^{-1}$ at $x = 6 \text{ \AA}$). This is a small variation in comparison with the changes in water energetics that dominate the thermodynamic signature. Decreasing the cavity base from 15 \AA to 12 \AA causes an unfavorable change in enthalpy of binding of ca. $+13 \text{ kcal}\cdot\text{mol}^{-1}$ and a favorable change in entropy of binding of ca. $-2 \text{ kcal}\cdot\text{mol}^{-1}$ because a greater number of cavity water molecules are displaced. However, decreasing the cavity base from $x = 12 \text{ \AA}$ to $x = 9 \text{ \AA}$ causes a favorable change in the enthalpy of binding by ca. $-11 \text{ kcal}\cdot\text{mol}^{-1}$ whereas the entropy of binding is left unchanged. This occurs because guest binding now displaces a mixture of stable and unstable water molecules. Finally, after decreasing the cavity base from $x = 9 \text{ \AA}$ to $x = 6 \text{ \AA}$ there is now a favorable change in the enthalpic contribution of ca. $-20 \text{ kcal}\cdot\text{mol}^{-1}$ that is offset by an unfavorable change in binding entropy of ca. $+7 \text{ kcal}\cdot\text{mol}^{-1}$. As shown in Figure 11 these large variations in thermodynamic signature reflect well the changes in water energetics between the unbound and bound host structures. Strikingly, the changes in enthalpy or entropy do not correlate with the number of cavity water molecules displaced; at $x = 9 \text{ \AA}$, about 25 water molecules are displaced upon guest binding but at a lower energetic cost than for $x = 12 \text{ \AA}$, where only about 20 water molecules are displaced upon guest binding. Remarkably, at $x = 6 \text{ \AA}$, displacing the cavity water molecules has both an unfavorable enthalpic and entropic component. This is because the few water molecules left at the bottom of the cavity have greater entropy than in bulk (Figure 9, third row). The present results provide strong evidence that there is no general thermodynamic signature associated with hydrophobic host/guest association in aqueous conditions; rather the details of the energetics of each displaced cavity water molecule dictate the magnitude and sign of the entropic and enthalpic components of cavity dehydration.

4. Conclusions

In summary the cell theory model of liquid water is shown to yield excess entropies in good agreement with reference thermodynamic integration calculations for clusters of confined water molecules. In general a better agreement is observed for increased confinement. The main advantage of the cell theory model is that entropies are readily obtained from a single simulation whereas calculation of the same quantity with thermodynamic integration requires the design of suitable pathways to converge free energy gradients and changes in total potential energy. In addition, it is straightforward to spatially localize within the grid cell theory framework excess enthalpic and entropic components. This was used here to perform detailed graphical analyses of the contribution of water energetics to the thermodynamic signature of non-polar host guest association in model cavities. The main results are that the favorable entropic contribution to the free energy of guest binding upon displacement of water molecules from a non-polar cavity is due to an increase in the number of hydrogen-bonding partners in solution rather than a change in the width of energy minima. Additionally, depending on the interplay between cavity geometry and polarity, guest binding can cause a local or long-ranged perturbation in interfacial host water structure. Finally, the energetics of cavity water displacement are strongly influenced by the cavity polarity and geometry. Thus there is no unique thermodynamic signature associated with non-polar host-guest association. The present results warrant further GCT analyses of more complex models of protein-ligand complexes to elucidate the thermodynamic or kinetic profiles of protein-ligand association. To this end it is desirable to extend the present work by generalizing CT models or coupling them with quasi-harmonic or mutual information expansion methodologies that yield conformational entropies for flexible hosts and guests.⁶⁸⁻⁶⁹

ASSOCIATED CONTENT

AUTHOR INFORMATION

Corresponding Author

* E-mail: mail@julienmichel.net

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

JM is supported by a University Research Fellowship from the Royal Society. RHH is supported by BBSRC Grant BB/K001558/1. This research was also supported by EPSRC through an award of a CASE studentship to GG and by a Royal Society research equipment grant (RG110450). The research leading to these results has also received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 336289 to JM.

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENTS

Gratitude is expressed to Dr Christopher Woods for assistance with scientific computations and for his continuing efforts to develop the Sire software library.

REFERENCES

1. Ball, P. *Nature* **2008**, 452 (7185), 291-292.
2. Hummer, G. *Nat. Chem.* **2010**, 2 (11), 906-907.
3. Setny, P.; Baron, R.; Kekenos-Huskey, P. M.; McCammon, J. A.; Dzubiella, J. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, 110 (4), 1197-1202.

4. Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2009**, *131* (42), 15403-15411.
5. Lazaridis, T. *J. Phys. Chem. B* **1998**, *102* (18), 3531-3541.
6. Lazaridis, T. *J. Phys. Chem. B* **1998**, *102* (18), 3542-3550.
7. Lazaridis, T. *J. Phys. Chem. B* **2000**, *104* (20), 4964-4979.
8. Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (3), 808-813.
9. Nguyen, C. N.; Young, T. K.; Gilson, M. K. *J. Chem. Phys.* **2012**, *137* (4), 044101.
10. SZMAP, version 1.1.1, OpenEye Scientific Software, Inc., Santa Fe, NM, USA, www.eyesopen.com, 2013, accessed 10/06/2014.
11. Frolov, A. I.; Ratkova, E. L.; Palmer, D. S.; Fedorov, M. V. *J. Phys. Chem. B* **2011**, *115* (19), 6011-6022.
12. Sindhikara, D. J.; Hirata, F. *J. Phys. Chem. B* **2013**, *117* (22), 6718-6723.
13. Truchon, J.-F.; Pettitt, B. M.; Labute, P. *J. Chem. Theory Comput.* **2014**, *10* (3), 934-941.
14. Zhou, S.; Rogers, K. E.; de Oliveira, C. A. F.; Baron, R.; Cheng, L.-T.; Dzubiella, J.; Li, B.; McCammon, J. A. *J. Chem. Theory Comput.* **2013**, *9* (9), 4195-4204.
15. Huber, R. G.; Fuchs, J. E.; von Grafenstein, S.; Laner, M.; Wallnoefer, H. G.; Abdelkader, N.; Kroemer, R. T.; Liedl, K. R. *J. Phys. Chem. B* **2013**, *117* (21), 6466-6472.
16. Barillari, C.; Taylor, J.; Viner, R.; Essex, J. W. *J. Am. Chem. Soc.* **2007**, *129* (9), 2577-2587.
17. Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2009**, *113* (40), 13337-13346.
18. Yu, H.; Rick, S. W. *J. Phys. Chem. B* **2010**, *114* (35), 11552-11560.
19. Bodnarchuk, M.; Viner, R.; Michel, J.; Essex, J. W. *J. Chem. Inf. Model.* **2014**, Article ASAP, DOI: 10.1021/ci400674k.
20. Ross, G. A.; Morris, G. M.; Biggin, P. C. *PLoS ONE* **2012**, *7* (3), e32036.
21. Baron, R.; Setny, P.; McCammon, J. A. *J. Am. Chem. Soc.* **2010**, *132* (34), 12091-12097.
22. Snyder, P. W.; Mecinovic, J.; Moustakas, D. T.; Thomas, S. W., III; Harder, M.; Mack, E. T.; Lockett, M. R.; Heroux, A.; Sherman, W.; Whitesides, G. M. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (44), 17889-17894.
23. Graziano, G. *Chem. Phys. Lett.* **2012**, *533*, 95-99.
24. Irudayam, S. J.; Henchman, R. H. *J. Phys.: Condens. Matter* **2010**, *22* (28), 284108.
25. Michel, J. *Phys. Chem. Chem. Phys.* **2014**, *16*, 4465-4477.
26. Gerogiokas, G.; Calabro, G.; Henchman, R. H.; Southey, M. W. Y.; Law, R. J.; Michel, J. *J. Chem. Theory Comput.* **2014**, *10* (1), 35-48.
27. Henchman, R. H. *J. Chem. Phys.* **2007**, *126* (6), 064504.
28. Hua, L.; Zangi, R.; Berne, B. J. *J. Phys. Chem. C* **2009**, *113* (13), 5244-5253.
29. Berne, B. J.; Weeks, J. D.; Zhou, R., Dewetting and Hydrophobic Interaction in Physical and Biological Systems. In *Annu. Rev. Phys. Chem.*, 2009; Vol. 60, pp 85-103.
30. Wang, L.; Friesner, R. A.; Berne, B. J. *J. Phys. Chem. B* **2010**, *114* (21), 7294-7301.
31. Wang, L.; Friesner, R. A.; Berne, B. J. *Faraday Discuss.* **2010**, *146*, 247-262.
32. McQuarrie, D. A. *Statistical Mechanics*. 1st ed ; Harper Collins: New York, NY, 1976.
33. Pauling, L. *J. Am. Chem. Soc.* **1935**, *57*, 2680-2684.
34. Irudayam, S. J.; Henchman, R. H. *Mol. Phys.* **2011**, *109* (1), 37-48.
35. Irudayam, S. J.; Henchman, R. H. *J. Phys. Chem. B* **2009**, *113* (17), 5871-5884.

36. Horn, H. W.; Swope, W. C.; Pitner, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120* (20), 9665-9678.
37. Case, D.; Darden, T. A.; Cheatham, T. E.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Crowley, M.; Walker, R.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P., *Amber 11*. University of California: San Francisco, 2010.
38. Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. *J. Am. Chem. Soc.* **1984**, *106* (22), 6638-6646.
39. Uberuaga, B.; Anghel, M.; Voter, A. *J. Chem. Phys.* **2004**, *120* (14), 6363-6374.
40. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81* (8), 3684-3690.
41. Miyamoto, S.; Kollman, P. *J. Comput. Chem.* **1992**, *13* (8), 952-962.
42. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. *J. Comput. Phys.* **1977**, *23* (3), 327-341.
43. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98* (12), 10089-10092.
44. Woods, C. J.; Michel, J. *Sire Molecular Simulation Framework*, Revision 2014 & 2373; UK: 2013.
45. Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D.; Tye, T.; Houston, M.; Stich, T.; Klein, C.; Shirts, M. R.; Pande, V. S. *J. Chem. Theory Comput.* **2013**, *9* (1), 461-469.
46. Tironi, I. G.; Sperb, R.; Smith, P. E.; Vangunsteren, W. F. *J. Chem. Phys.* **1995**, *102* (13), 5451-5459.
47. Andersen, H. C. *J. Chem. Phys.* **1980**, *72* (4), 2384-2393.
48. Mezei, M. *J. Chem. Phys.* **1987**, *86* (12), 7084-7088.
49. Zacharias, M.; Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994**, *100* (12), 9025-9031.
50. Michel, J.; Verdonk, M. L.; Essex, J. W. *J. Chem. Theory Comput.* **2007**, *3* (5), 1645-1655.
51. Shyu, C.; Ytreberg, F. M. *J. Comput. Chem.* **2009**, *30* (14), 2297-2304.
52. Latimer, W. M.; Pitzer, K. S.; Slansky, C. M. *J. Chem. Phys.* **1939**, *7* (2), 108-111.
53. Conway, B. E. *J. Solution Chem.* **1978**, *7* (10), 721-770.
54. Hummer, G.; Pratt, L. R.; Garcia, A. E. *J. Phys. Chem.* **1996**, *100* (4), 1206-1215.
55. LyndenBell, R. M.; Rasaiah, J. C. *J. Chem. Phys.* **1997**, *107* (6), 1981-1991.
56. Grossfield, A. *J. Chem. Phys.* **2005**, *122* (2), 024506.
57. Mobley, D. L.; Barber, A. E., II; Fennell, C. J.; Dill, K. A. *J. Phys. Chem. B* **2008**, *112* (8), 2405-2414.
58. Mukhopadhyay, A.; Fenley, A. T.; Tolokh, I. S.; Onufriev, A. V. *J. Phys. Chem. B* **2012**, *116* (32), 9776-9783.
59. Qvist, J.; Davidovic, M.; Hamelberg, D.; Halle, B. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (17), 6296-6301.
60. Young, T.; Hua, L.; Huang, X.; Abel, R.; Friesner, R.; Berne, B. J. *Proteins: Struct., Funct., Bioinf.* **2010**, *78* (8), 1856-1869.
61. Matthews, B. W.; Liu, L. *Protein Sci.* **2009**, *18* (3), 494-502.
62. Syme, N. R.; Dennis, C.; Bronowska, A.; Paesen, G. C.; Homans, S. W. *J. Am. Chem. Soc.* **2010**, *132* (25), 8682-8689.
63. Dunitz, J. D. *Science* **1994**, *264* (5159), 670-670.

- 64. Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W. M.; Taylor, R. D.; Taylor, R. *J. Med. Chem.* **2005**, *48* (20), 6504-6515.
- 65. Li, Z.; Lazaridis, T. *J. Am. Chem. Soc.* **2003**, *125* (22), 6636-6637.
- 66. Li, Z.; Lazaridis, T. *J. Phys. Chem. B* **2005**, *109* (1), 662-670.
- 67. Li, Z.; Lazaridis, T. *J. Phys. Chem. B* **2006**, *110* (3), 1464-1475.
- 68. Killian, B. J.; Kravitz, J. Y.; Gilson, M. K. *J. Chem. Phys.* **2007**, *127* (2), 024107.
- 69. Baron, R.; Huenenberger, P. H.; McCammon, J. A. *J. Chem. Theory Comput.* **2009**, *5* (12), 3150-3160.
- 70. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics Modell.* **1996**, *14* (1), 33-38.

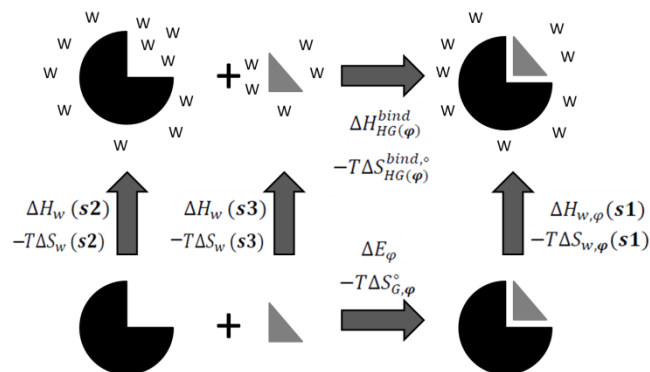


Figure 1. Thermodynamic cycle for the computation of enthalpies and entropies of binding of rigid guest G with rigid host H . The symbols are defined in section 2.1.

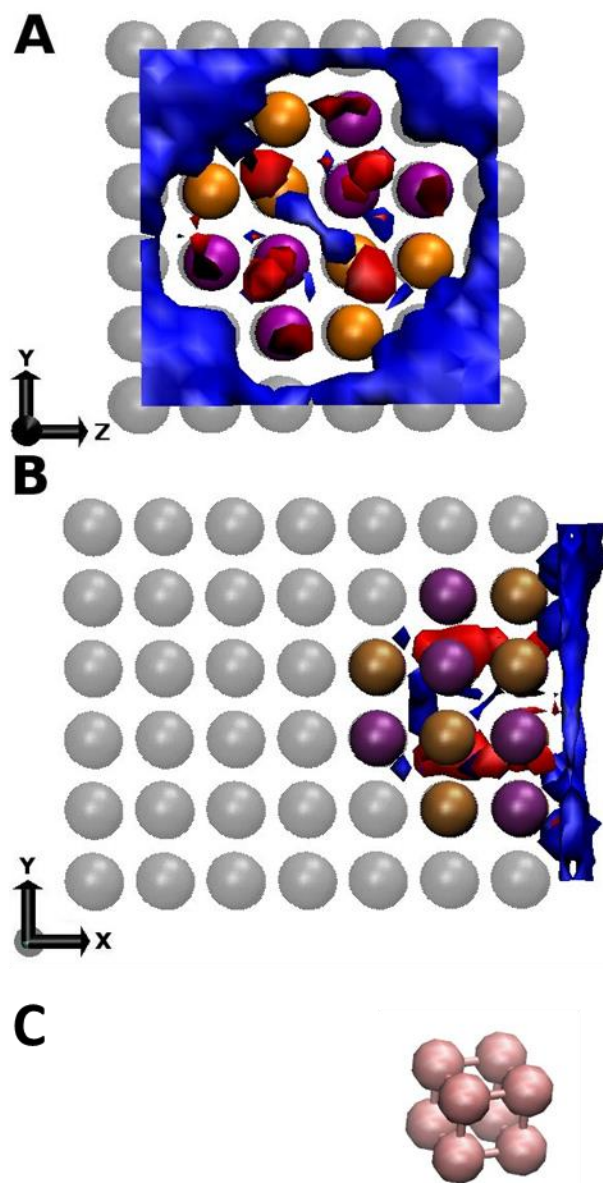


Figure 2. Hydration thermodynamics of a model host cavity. A) Top view in the zy plane. B) Side view in the xy plane. Host sites are spaced every 3.8 Å. The solvent accessible volume of the cavity is approximately 350 Å³. Sites bearing a partial positive or negative charge ($|q| = 0.55$ e) are represented as orange or purple spheres respectively. Other sites in the host are uncharged and represented as transparent black spheres. Isocontours of $(\Delta G_w^n(\mathbf{s}))$ the normalized free energy of water have been drawn in red (-2.5 kcal•mol⁻¹) and blue (0.5 kcal•mol⁻¹). C) The guest

is modelled as a set of eight Lennard-Jones sites sized to approximately fill up the host cavity.

Figure prepared with the software VMD.⁷⁰

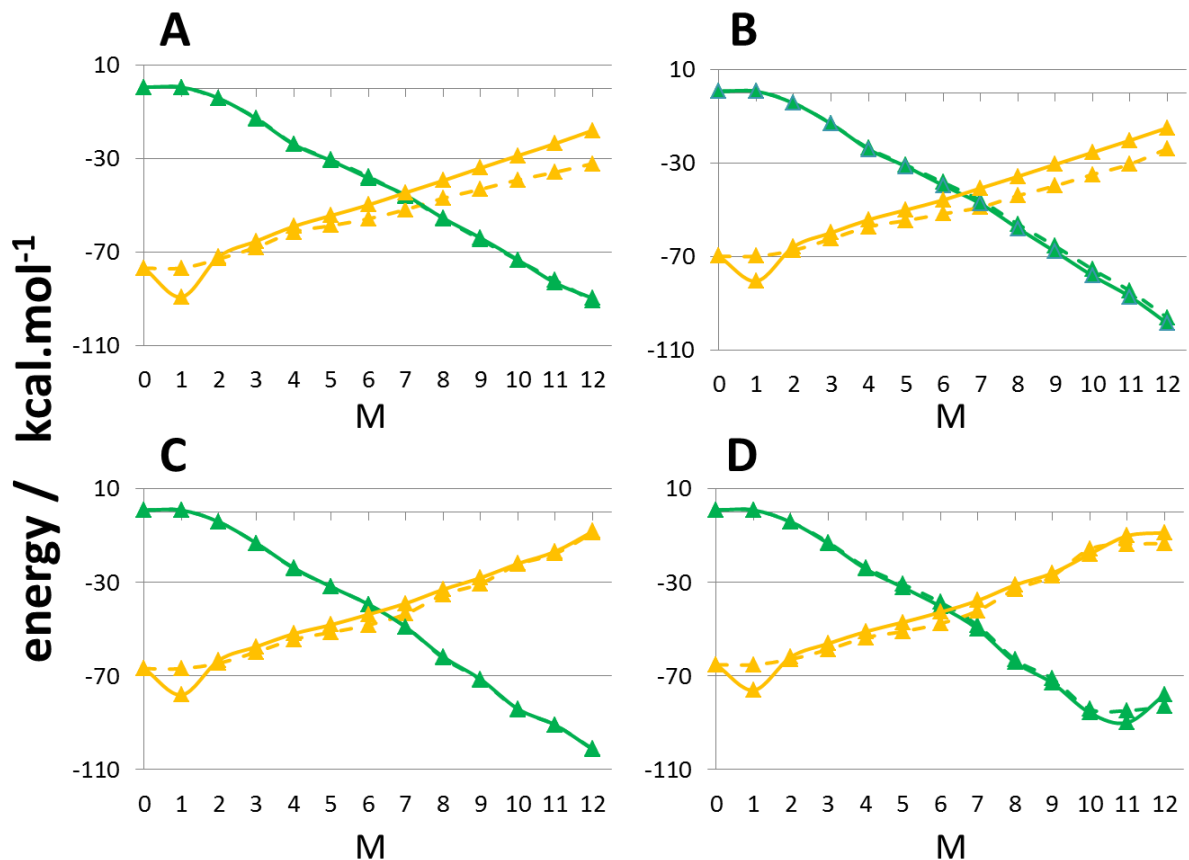


Figure 3. The excess internal energy $\Delta U_{excess}^{N,M}$ (green) and temperature-entropy $-T\Delta S_{excess}^{N,M}$ (orange) of clusters of $N + M = 12$ water molecules under varying degree of confinement. N is the number of non-interacting water molecules, M is the number of interacting water molecules. A) $D=5.0$ Å, B) $D=3.5$ Å, C) $D=3.0$ Å, D) $D=2.75$ Å. The solid lines are for the cell theory results; the dashed-lines are for the thermodynamic integration results.

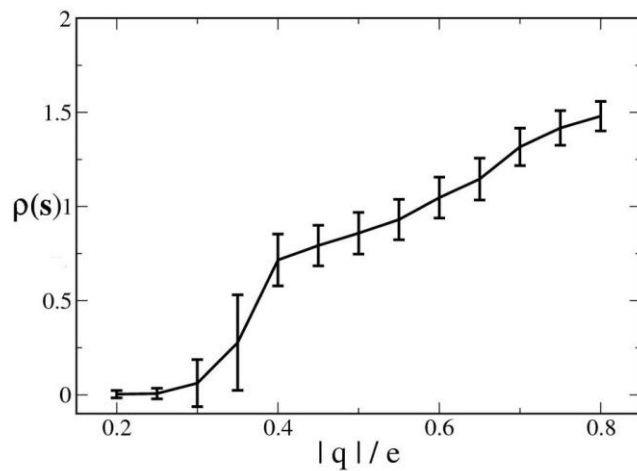


Figure 4. The effect of polarity on the solvent density of a model cavity. The y axis shows the cavity density relative to bulk water, the x axis the magnitude of partial charge on the host polar sites. The error bars show one standard-deviation of the mean density. See figure 2 for a representation of the cavity region s .

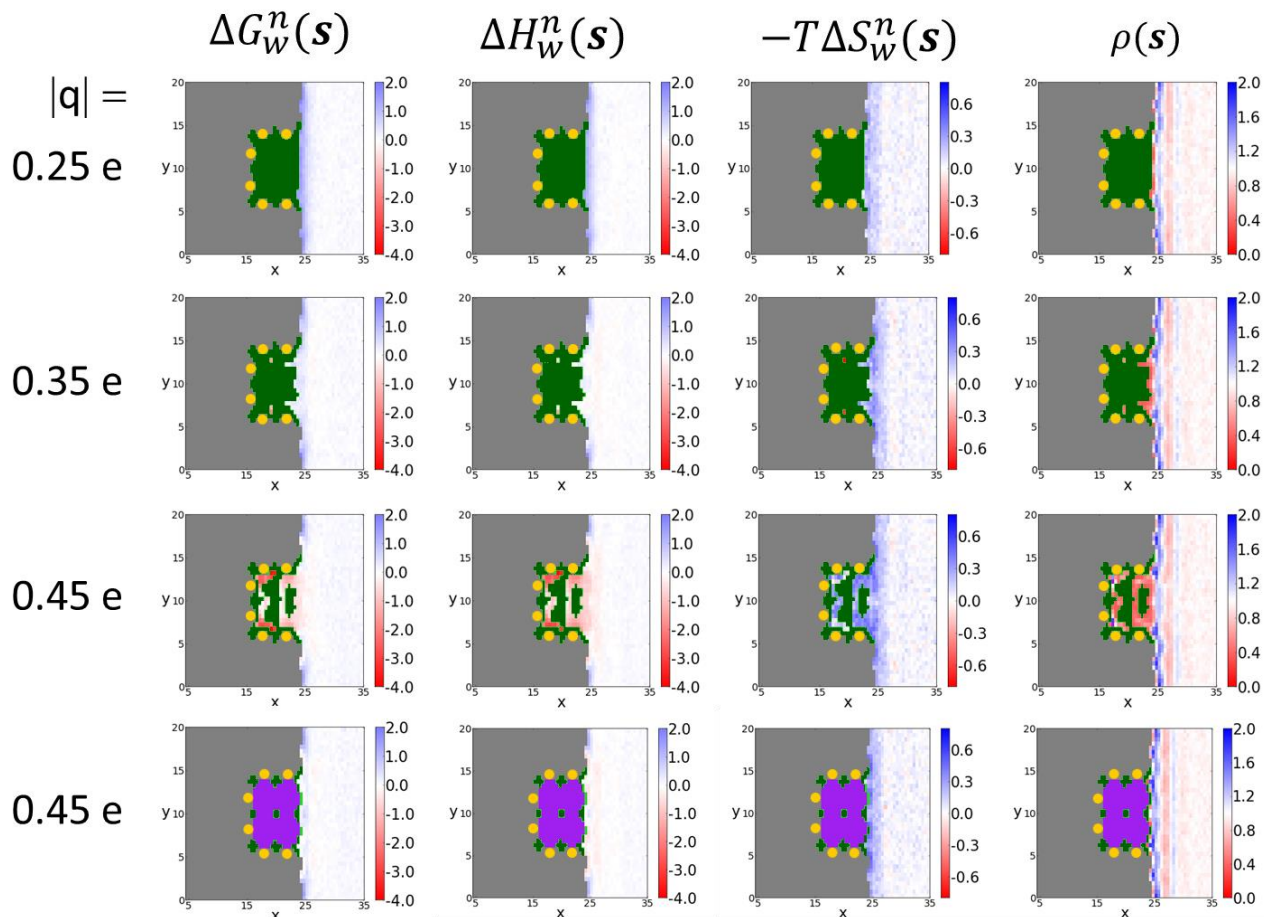


Figure 5. The effect of cavity polarity on the energetics of interfacial water molecules. Water properties have been projected in the xy plane. From left to right the panels depict normalized water free energies $\Delta G_w^n(s)$, normalized water enthalpies $\Delta H_w^n(s)$, normalized water entropies $-T\Delta S_w^n(s)$ ($\text{kcal}\cdot\text{mol}^{-1}$) and relative water density $\rho(s)$ (red-white-blue heatmap). From top to bottom the magnitude of the partial charges $|q|$ on the host cavity polar sites increases from 0.25 to 0.45 e. The volume of space displaced by host particles is depicted in gray and by guest particles in purple. The orange circles indicate the position of the host polar particles. The color green denotes solvent-accessible regions of space that have less than 1% bulk solvent density. For clarity host particles in the z planes above and below the cavity have been hidden. The x and

y axes units are in Å and are numbered such that (0,0) is the coordinate of the host site with the lowest x and y values.

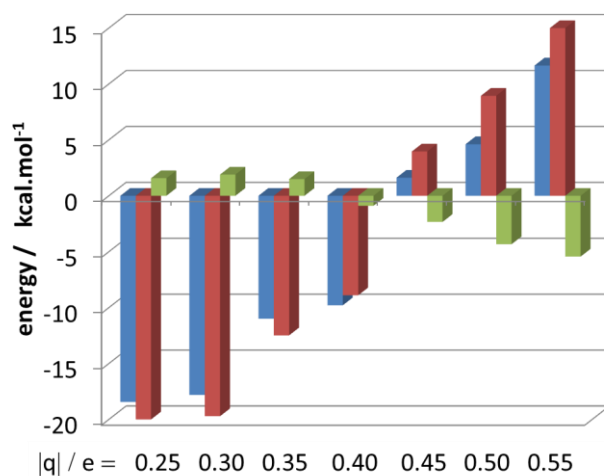


Figure 6. The thermodynamic signature of host/guest binding as a function of the host cavity polarity. Blue: free energy of binding $\Delta G_{HG(\varphi)}^{bind,\circ}$. Red: enthalpy of binding $\Delta H_{HG(\varphi)}^{bind}$. Green: entropy of binding $-T\Delta S_{HG(\varphi)}^{bind,\circ}$.

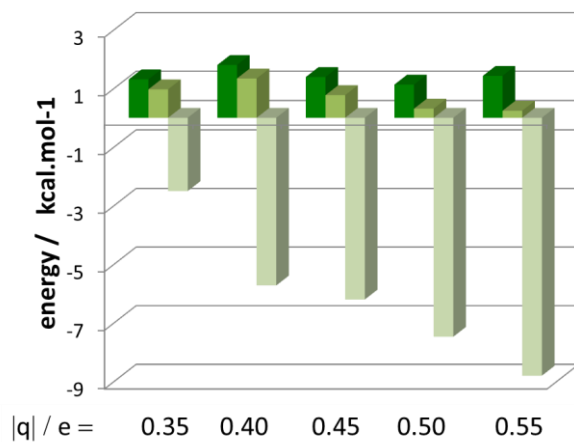


Figure 7. The influence of the cavity polarity on the changes in the entropic components of water upon guest binding. Dark green: changes in vibrational entropy $-T[\Delta S_{w,\varphi}^{vib}(\mathbf{s1}) - \Delta S_w^{vib}(\mathbf{s2})]$. Green: changes in librational entropy $-T[\Delta S_{w,\varphi}^{lib}(\mathbf{s1}) - \Delta S_w^{lib}(\mathbf{s2})]$. Light green: changes in orientational entropy $-T[\Delta S_{w,\varphi}^{ori}(\mathbf{s1}) - \Delta S_w^{ori}(\mathbf{s2})]$.

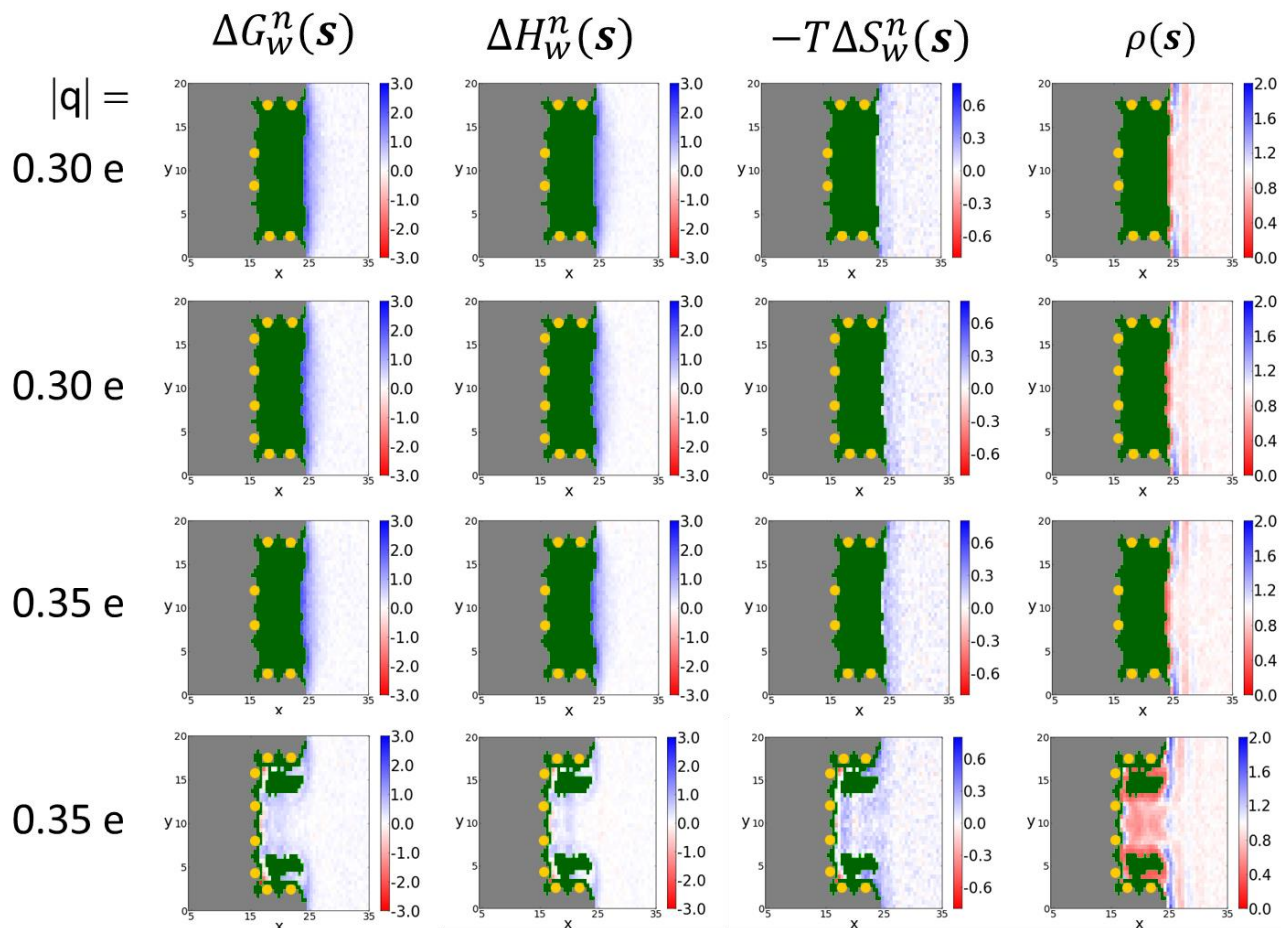


Figure 8. The effect of cavity width on the energetics of interfacial water. In comparison with Figure 5, the cavity has been enlarged by $\pm r_{eq}$ Å along the y and z planes. From left to right the panels depict normalized water free energies $\Delta G_w^n(s)$, normalized water enthalpies $\Delta H_w^n(s)$, normalized water entropies $-T\Delta S_w^n(s)$ (kcal•mol⁻¹) and relative water density $\rho(s)$. From top to bottom panels: $|q| = 0.30$ e; $|q| = 0.30$ e with additional polar sites; $|q| = 0.35$ e; $|q| = 0.35$ e with additional polar sites. Other symbols as in Figure 5.

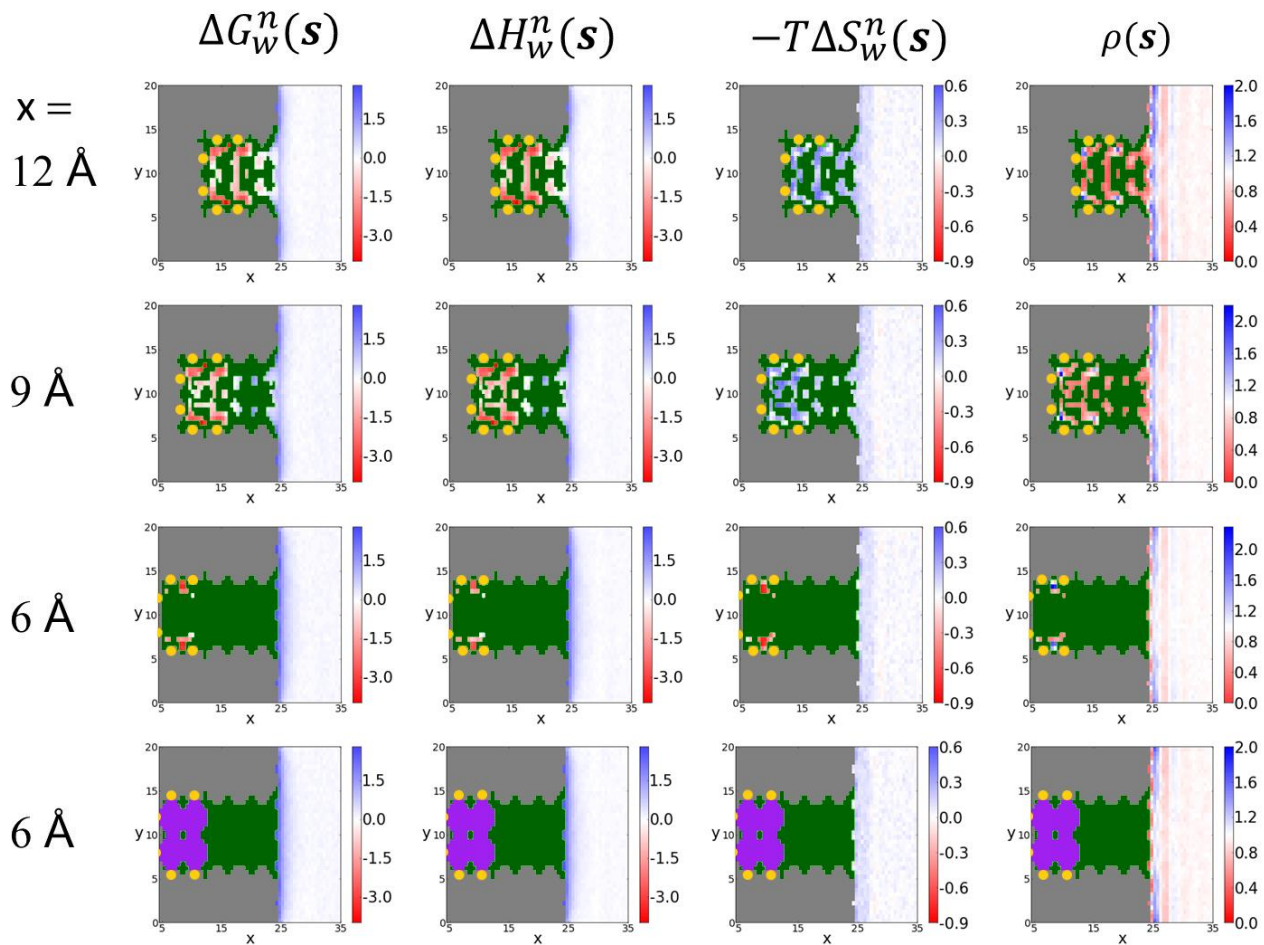


Figure 9. The effect of cavity depth on the energetics of interfacial water. From left to right the panels depict normalized water free energies $\Delta G_w^n(s)$, normalized water enthalpies $\Delta H_w^n(s)$, normalized water entropies $-T\Delta S_w^n(s)$ (kcal•mol⁻¹) and relative water density $\rho(s)$. From top to bottom panels the x coordinate of the base of the cavity decreases from 12 Å to 6 Å. $|q| = 0.45$ e for all cavities. Other symbols as in Figure 5.

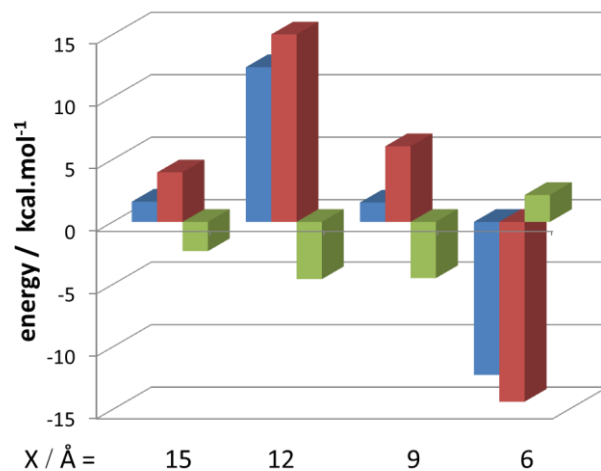


Figure 10. The thermodynamic signature of host/guest binding as a function of the host cavity depth. $|q| = 0.45$ e for all cavities. Blue: free energy of binding $\Delta G_{HG(\varphi)}^{bind, \circ}$. Red: enthalpy of binding $\Delta H_{HG(\varphi)}^{bind}$. Green: entropy of binding $-T\Delta S_{HG(\varphi)}^{bind, \circ}$.

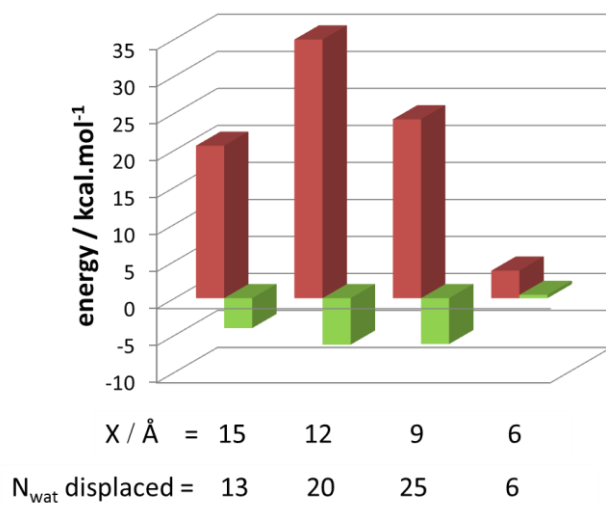


Figure 11. The contribution of cavity dehydration to the thermodynamics of host/guest binding as a function of the cavity depth. $|q| = 0.45$ e for all cavities. Red: change in water enthalpy $\Delta H_{w,\varphi}(\mathbf{s1}) - \Delta H_w(\mathbf{s2})$. Green: change in water entropy $-T[\Delta S_{w,\varphi}(\mathbf{s1}) - \Delta S_w(\mathbf{s2})]$.

Graphical TOC

